JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, D05106, doi:10.1029/2007JD009278, 2008

Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model

Thomas Reichler[1] and Junsu Kim[1]

Received 13 August 2007; revised 15 November 2007; accepted 12 December 2007; published 5 March 2008.

[1] Climate research relies on realistic atmospheric data over long periods of time. Global reanalyses or observations are commonly used for this type of work. However, the many problems associated with both the reanalyses and observations cast doubts on the reliability of such data for climate applications, and users often need to know how large the errors and uncertainties associated with the different data sets are. This paper is a systematic assessment of the errors and uncertainties contained in the time mean (1979–1999) of many different climate quantities taken from a variety of global data sets, including four popular reanalyses, the output of the climate model developed at the Geophysical Fluid Dynamics Laboratory (GFDL), and a wide range of observations. We find that the ability of reanalyses to reproduce the observed climate mean state varies widely, with radiative quantities exhibiting the largest discrepancies. The different reanalysis products share many common errors, but overall the European Centre for Medium-Range Weather Forecasts 40-year reanalysis (ERA-40) matches best the observations. Interestingly, the climate model reproduces the observed climate mean state of certain quantities more faithfully than the reanalyses. This indicates that modern models have reached a high level of realism in their mean state and that care must be taken when reanalyses are used to validate models. A particular concern of this paper is the time mean uncertainty associated with specific observation-based atmospheric quantities. Observational uncertainties are estimated from the difference amongst alternative data sets for the same quantity. We show that for most quantities the observational uncertainty is smaller than the error of the reanalyses or the model. However, there are some notable exceptions. In particular, for the surface fluxes of heat, momentum, and radiation the observational uncertainties can be as large as the errors seen in the reanalyses or the model. The investigation of uncertainties in upper atmospheric quantities is restricted to reanalysis and model data, since no appropriate observations are available. In this case, the reanalyses uncertainties are generally smaller than the model errors, except for quantities which describe the meridional component of the atmospheric circulation.

Citation: Reichler, T., and J. Kim (2008), Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model, *J. Geophys. Res.*, *113*, D05106, doi:10.1029/2007JD009278.

1. Introduction

[2] Multidecadal data sets that accurately describe the state of the global atmosphere are a common need in climate research. For example, testing and validating climate models requires access to global climate data for many different quantities. Global atmospheric reanalyses are often used for this type of work, in particular when real observations are unavailable, or when continuous spatial and temporal coverage is mandatory. Reanalyses are the product of numerical data assimilation, composing past observations of various sources to a dynamically consistent three-dimensional picture of the atmosphere at specific time intervals. A variety of reanalyses from different centers around the world are available to the research community. From the user perspective, this choice raises the following questions: What are the specific strengths and weaknesses of the individual products in depicting the real state of the atmosphere, and which product should be used for a specific application?

[3] Although reanalyses have substantially improved over time, mainly owing to the use of more realistic models and the inclusion of additional observations [*Kanamitsu et al.*, 2002; *Uppala et al.*, 2005], there are still a variety of limitations associated with reanalyses, raising doubts about their usefulness for climate studies. During the assimilation process, multiple sources of errors can affect the quality of the reanalyses: The general difficulty in combining heterogeneous observations onto a regular grid, uncertainties regarding the assimilation model, and the quality and distribution of the underlying observations [*Bengtsson et*

[1]Department of Meteorology, University of Utah, Salt Lake City, Utah, USA.

0148-0227/08/2007JD009278$09.00

Table 1. Reanalyses and Model Data Considered in This Study[a]

Name	Acronym	Reference
National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis	NNR	*Kalnay et al.* [1996]
NCEP/Department of Energy (DOE) reanalysis	NDR	*Kanamitsu et al.* [2002]
European Centre for Medium-Range Weather Forecasts 40-year reanalysis	ERA	*Uppala et al.* [2005]
Japanese 25-year reanalysis	JRA	*Onogi et al.* [2007]
GFDL CM2.1 coupled climate model (one member of 20C3M ensemble)	GFD	*Delworth et al.* [2006]

[a]The base period for all reanalyses and model data was 1979–1999.

al., 2004]. The resulting problems have been discussed in a number of studies. For example, *Marshall* [2003], *Renwick* [2004], *Bengtsson et al.* [2004], and *Trenberth et al.* [2001] all showed that reanalyses are of limited usefulness for climate studies because of spurious trends introduced by changes in the observation system. Moreover, different reanalyses lead sometimes to inconsistent results [*Betts et al.*, 2006; *Bromwich and Fogt*, 2004; *Marshall*, 2002; *Newman et al.*, 2000; *Sterl*, 2004], indicating that at least one of the products is flawed in some aspect.

[4] For a limited number of quantities, real observations are also available. However, like reanalyses, observations are not perfect. For example, most observations suffer instrument and processing problems. In addition, observation-based global data sets are often derived from a few localized measurements, and statistical approximations are used to fill the gaps. Local observations, however, are not always good representatives for their surroundings. Also, some quantities cannot be directly measured, but are instead derived from other observations, multiplying the uncertainties in the final result. Another severe limitation of many observations is that their time period covered is often too short to be useful for climate applications.

[5] The present study was motivated by the need to better understand the uncertainties associated with global reanalyses and observations in the context of climate model validation. Climate models are continuously improved. Most currently available reanalyses, on the other hand, depend on systems designed many years ago, when data assimilation and atmospheric modeling was less advanced. This discrepancy may lead to models gradually outpacing the reanalyses. As pointed out before, global observations are also problematic, and one often wishes to better understand how large their uncertainties are. In the present study, we address these issues from one specific angle: We focus exclusively on the time mean state of climate. In particular, we investigate the uncertainties in the climatologies of many quantities, derived from a variety of products, including a wide range of modern observations, four different reanalysis data sets, and the output of one state-of-the-art climate model.

[6] This paper is structured as follows: In sections 2 and 3, we describe the data and the methods. In section 4, we present an illustrating example, and in section 5, we show the systematic errors of the reanalyses. The main results of this study (the errors and uncertainties of observations, reanalyses, and model) are presented in section 6. Section 7 compares the uncertainties seen in the individual products. Section 8 attempts to quantify the overall errors seen in the different reanalyses. Section 9 contains a summary of our findings and some conclusions.

2. Data

[7] In this study we compared climatologies of four different reanalyses (Table 1). First, we used the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research reanalysis [*Kalnay et al.*, 1996] (NNR), which is the first atmospheric reanalysis performed over a long period of time. NNR is still widely used in climate and atmospheric studies. Second, we investigated the NCEP/Department of Energy reanalysis [*Kanamitsu et al.*, 2002] (NDR), which is similar to NNR, except for some bug fixes and minor improvements. NNR and NDR both have a base resolution of T62 with 28 vertical levels, and are considered first generation products. Third, we included the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year reanalysis [*Uppala et al.*, 2005] (ERA), which is considered to be a second generation reanalyses. ERA attempts to improve many of the problems encountered with the previous reanalyses, and it features a higher horizontal and vertical resolution (T159, 60 levels) than NNR and NDR. Fourth, we used the Japanese 25-year reanalysis [*Onogi et al.*, 2007] (JRA), which represents the latest generation of reanalyses. JRA has a spectral resolution of T106 with 40 vertical levels, and assimilates more observations than NNR, NDR, and ERA. Another major reanalysis project, the ERA-15 [*Gibson et al.*, 1997], produced by ECMWF, was not considered in this study since it covers a shorter period of time and since it was entirely superseded by the more modern ERA.

[8] We also included the output of one state-of-the-art coupled climate model in our intercomparison. This provides a benchmark against which the errors seen in the reanalyses can be compared. The selected model output is one member integration of the "climate of the twentieth century" simulation by the CM2.1 model [*Delworth et al.*, 2006], developed at the Geophysical Fluid Dynamics Laboratory (GFDL). This data set will be denoted by GFD.

[9] Table 2 provides an overview of the climate quantities included in our intercomparison and the observational data sets used. Most quantities were horizontal fields of global extent. Others only covered the global oceans, and still others were zonal mean latitude-height cross sections, ranging from 90°S to 90°N and from 1000 and 100 hPa. On the basis of the availability of validating observations, we divided the quantities into two groups. Quantities belonging to the first group were validated against real observations. According to the predominant nature of its

Table 2. Climate Quantities Analyzed in This Study[a]

Quantity v	Domain	Acronym	Units	Validating Observations O_{1-5}
Physics				
Surface air temperature	global	TAS	K	CRU, ICOADS, NOAA
Zonal/meridional surface wind stress	ocean	TAUU, TAUV	10^{-2} N m^{-2}	GSSTF2, ICOADS
Sea level pressure	ocean	PSL	hPa	ERSLP, HADSLP, ICOADS
Surface sensible/latent heat fluxes	ocean	HFSS, HFLS	W m^{-2}	GSSTF2, HOAPS2, ICOADS, JOFURO, OAFLUX
Total cloudiness	global	CLT	%	CERES, ISCCP
Surface radiation (up/down, shortwave/longwave)	global	RSDS, RSUS, RLDS, RLUS	W m^{-2}	BSRN, CERES, GEBA, ISCCP
TOA outgoing shortwave radiation	global	RSUT	W m^{-2}	CERES, ERBE, ISCCP
TOA outgoing longwave radiation	global	RLUT	W m^{-2}	CERES, ERBE, ISCCP, NOAA
TOA cloud radiative forcing	global	CFLT, CFST	W m^{-2}	CERES, ERBE, ISCCP
Precipitation	global	PR	mm/d	CMAP, GPCP
Precipitable water	global	PRW	mm	HOAPS2, NVAP
Air temperature	zonal mean	TA	K	AIRS
Dynamics				
Specific humidity	zonal mean	HUS	g/kg	ERA
Zonal/meridional wind 200 hPa	global	U200, V200	m/s	ERA
Stream function 200 hPa	global	χ200	10^6 m^2 s^{-1}	ERA
Velocity potential 200 hPa	global	ψ200	10^6 m^2 s^{-1}	ERA
Temperature 200 hPa	global	T200	K	ERA
Geopotential 500 hPa	global	Z500	gpm	ERA
Stationary waves 500 hPa	global	SW500	gpm	ERA
Zonal/meridional wind 850 hPa	global	U850, V850	m/s	ERA
Zonal mean zonal/meridional wind	zonal mean	UA, VA	m/s	ERA
Mean meridional mass stream function	zonal mean	MMC	10^9 kg/s	ERA

[a]Quantities listed as "zonal mean" are latitude (90°S–90°N)-height distributions of zonal averages on 12 atmospheric pressure levels between 1000 and 100 hPa. Because of limitations of NNR and NDR, specific humidity (HUS) was only calculated between 1000 and 300 hPa. Quantities listed as "global" or "ocean" are global single-level fields for the respective regions. SW500, the stationary waves at 500 hPa, are calculated from 500 hPa geopotential heights (Z500) minus zonal mean of Z500. The mean meridional stream function (MMC) was calculated as described by *Lu et al.* [2007]. Cloud radiative forcing for longwave and shortwave (CFLT, CFST) is defined as the difference between total and clear sky forcing at the top of the atmosphere. See Table 3 for more information on the validation data sets.

members, this group will be simply referred to as "physics" group. The second group is mostly composed of upper air dynamical quantities, for which no real observations were available. We will refer to this group as "dynamics" group.

[10] The "physics" group encompasses 18 quantities, most of which characterize the thermodynamic state of the atmosphere and the radiative energy passing through it. All quantities in this group were compared against observation-based data (hereafter simply "observations"), gathered by satellites, ground-based instruments, or a combination of both. The majority of the "physics" quantities were validated against multiple observations, which allowed us to estimate the reanalysis errors as well as the observational uncertainty associated with a particular quantity. In order to be useful as validation data, these observations had to meet certain requirements, such as sufficient global and temporal coverage and independence from the reanalyses. Unfortunately, it was not always possible to find adequate observations that fulfilled all those requirements, and as outlined below, certain compromises in terms of the validation procedure were sometimes unavoidable.

[11] We also included 13 "dynamics" quantities in our intercomparison, mostly describing the circulation of the free atmosphere. In this case, we chose to consider quantities that are commonly used for the testing of climate models. Because of the lack of true verifying observations, we compared this group of quantities against the ERA reanalysis. The decision to use ERA as reference was based upon the good performance of the ERA reanalysis for the "physics" quantities. As a disadvantage of this approach, it was impossible to verify ERA themselves or to obtain estimates of observational uncertainties.

3. Error Calculation

[12] This intercomparison is based on seasonal mean climatologies (DJF, MAM, JJA, and SON), which in most cases were derived from monthly mean data. (The BSRN climatology, the only exception, was derived from high-frequency (minute intervals) input data.) The climatological base period for most data (reanalyses, model, observations) was 1979–1999. However, because of restrictions imposed by the available data, some of the validating observations cover somewhat different base periods (Table 3). As explained below, we investigated the impact of taking climatologies over different time periods and found that the resulting differences were negligible.

[13] In our terminology, the resulting climatologies are $\overline{r}_{rqsn}$ and $\overline{o}_{oqsn}$. Here, the overbar indicates (time) averaging, and (r) indicates the reanalysis or the model used, (o) the validating observations, (q) the climate quantity, (s) the season, and (n) the grid point. If multiple observations for the same quantity were available, we interpolated the different observations to a common grid and took the mean over all available climatologies to form a new, best observational estimate, or in more mathematical terms, $o = \overline{o_o}^{o}$. If at certain grid points some observations were missing, we simply used the mean over the remaining observations. This averaging technique is similar to taking ensemble and multimodel averages in weather and climate forecasting,

Table 3. Observations and Base Periods for Calculating the Climatologies

Name	Reference	Period
AIRS	*Chahine et al.* [2006]	2002–2006
BSRN	*Ohmura et al.* [1998]	variable
CERES	*Wielicki et al.* [1996]	2000–2005
CMAP	*Xie and Arkin* [1996]	1979–1999
CRU	*Jones et al.* [1999]	1979–1999
ERA	*Uppala et al.* [2005]	1979–1999
ERBE	*Barkstrom* [1984]	1985–1989
ERSLP	*Smith and Reynolds* [2004]	1979–1999
GEBA	*Gilgen and Ohmura* [1999]	variable
GPCP	*Adler et al.* [2003]	1979–1999
GSSTF2	*Chou et al.* [2003]	1988–1999
HADSLP	*Allan and Ansell* [2006]	1979–1999
HOAPS2	*Grassl et al.* [2000]	1988–1999
ICOADS	*Woodruff et al.* [1987]	1979–1999
ISCCP	*Rossow et al.* [1996], *Zhang et al.* [2004]	1984–1999
JOFURO	*Kubota et al.* [2002]	1991–1998
NOAA	*Smith and Reynolds* [2005]	1979–1999
NOAA	*Liebmann and Smith* [1996]	1979–1999
NVAP	*Randel et al.* [1996]	1988–1999
OAFLUX	*Yu and Weller* [2007]	1981–1999

which, because of the cancellation of randomly distributed errors, is generally superior to using results from individual simulations [*Reichler and Kim*, 2008].

[14] Prior to calculating errors, we interpolated the reanalysis and model climatologies to the observational grid and calculated the difference $d = \overline{r} - \overline{o}$. Our most basic measure of error was the mean bias b, defined by $b = \Sigma_n w_n d_n$. Here, w_n are proper weights needed for the averaging, taking into account the different area and mass represented by the grid points as one goes poleward and upward, respectively.

[15] We also employed the normalized error variance E^2, defined by

$$E^2 = \Sigma_n w_n d_n^2 / \sigma_n^2. \quad (1)$$

Here, σ_n^2 is the interannual variance at grid point (n), derived from the observations. The normalization with the interannual variance was crucial since it nondimensionalized the results and helped to make the errors from different regions and quantities more comparable. (The advantages of normalized and nondimensional measures of errors in the context of model validation are discussed in more detail by *Watterson* [1996].) If multiple observations were available for the same quantity, σ_n^2 was derived from the data set with the longest available record. Prior to using σ_n^2, grid point values of zero were set to missing and the field was spatially smoothed using a running box average filter. We note that the normalization can also be performed using the spatial variance of the climatological mean. This alternative approach has been previously used by *Watterson and Dix* [2005] and *Watterson et al.* [1999].

[16] Lastly, in some of our analysis, we also utilized the ordinary (i.e., dimensional) error variance

$$e^2 = \Sigma_n w_n d_n^2. \quad (2)$$

[17] When displaying our results, we actually used E or e, which are the normalized and nonnormalized root mean square (RMS) errors, respectively. However, it is important to note that all statistical manipulations (e.g., averaging) were performed on the original squared quantities. Spatially, the various error measures were either calculated globally (GL) or for specific latitude bands corresponding to the Northern Extratropics (NH) (30–90°N), Tropics (TR) (30°S–30°N) and Southern Extratropics (SH) (90–30°S).

[18] The validation work involving observations from BSRN [*Ohmura et al.*, 1998] and GEBA [*Gilgen and Ohmura*, 1999] requires some additional explanation. BSRN and GEBA are high-quality surface radiation measurements taken at various distinct stations. Since the stations are spatially unevenly distributed, we only selected data from a subsample of stations in order to achieve an approximately even global coverage. The number and location of included stations differed from quantity to quantity. We validated reanalyses and model data against these station data by linearly interpolating the nearest grid points to the geographic station locations. When forming multiobservational means, data from GEBA and BSRN were excluded because of their poor spatial coverage.

4. Illustrating Example

[19] We start by illustrating our validation procedure using one specific, well observed quantity: the outgoing shortwave radiation at the top of the atmosphere (TOA) (RSUT). We first study the observational uncertainties associated with this quantity and then compare the annual mean RSUT climatologies from the three available observations (CERES, ERBE, ISCCP) (Figure 1, top). At first sight, the basic patterns displayed by these climatologies are similar. However, there exist subtle discrepancies, which become clearer when taking differences (Figure 1, bottom). For example, over most land surfaces, ISCCP exhibits smaller values than ERBE and CERES. These and other differences are related to variations in the underlying observation systems, retrieving algorithms, and base periods over which the climatologies were calculated (Table 3). If the observational uncertainty is taken to be the global RMS error e among the different observations, then this uncertainty is on the order of 10 W m^{-2} for RSUT.

[20] We now use the above three observational estimates of RSUT to compare the 1979–1999 annual mean climatologies of the four reanalyses and the model output (Figure 2). The first three columns display the differences between reanalyses and observations, or d_{ro}, with $r = 1 \ldots 5$ (including GFD), and $o = 1 \ldots 3$. All reanalysis products exhibit a general tendency of positive biases over the tropical and subtropical oceans, and predominantly negative biases over the remaining areas. Other common features are negative biases over the coastal upwelling regions and the Saharan dessert. These results are largely independent of our choice of validating observations. The resulting global RMS errors are smallest for JRA (16–17 W m^{-2}) and largest for NNR (20–28 W m^{-2}). Using an approximate mean value of 20 W m^{-2}, these RMS errors are about by a factor of two larger than the RMS error amongst the three observations. We conclude that in this case the

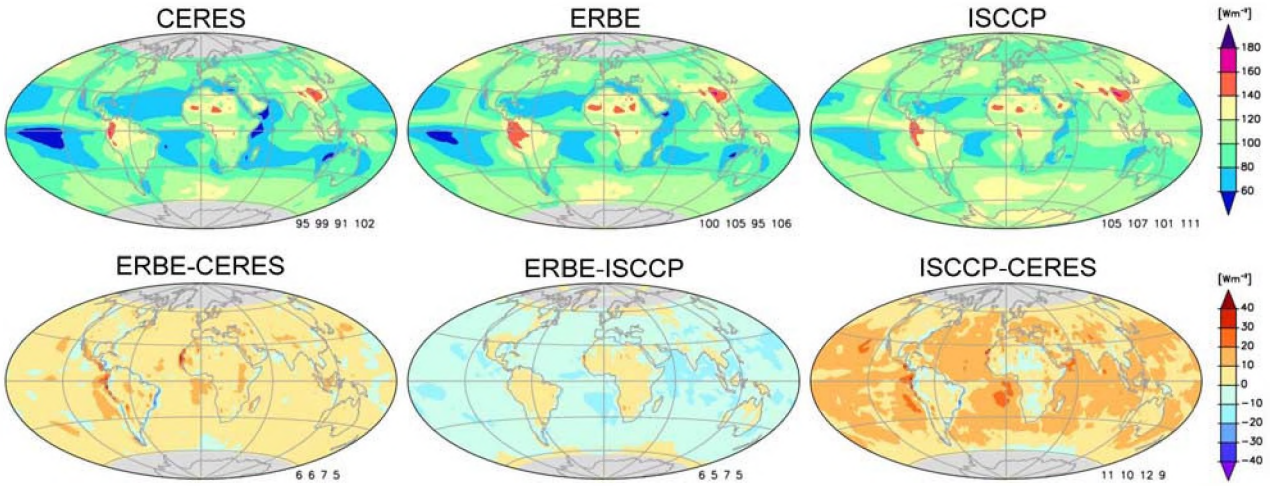

Figure 1. Observed annual mean climatology of outgoing shortwave radiation at the top of the atmosphere (RSUT). Shown are (top) full fields and (bottom) differences among the three observations. Numbers in the lower right corners indicate (top) spatial means and (bottom) RMS errors e for the four regions (GL, NH, TR, and SH). Grey shading indicates insufficient data. Units are W m^{-2}.

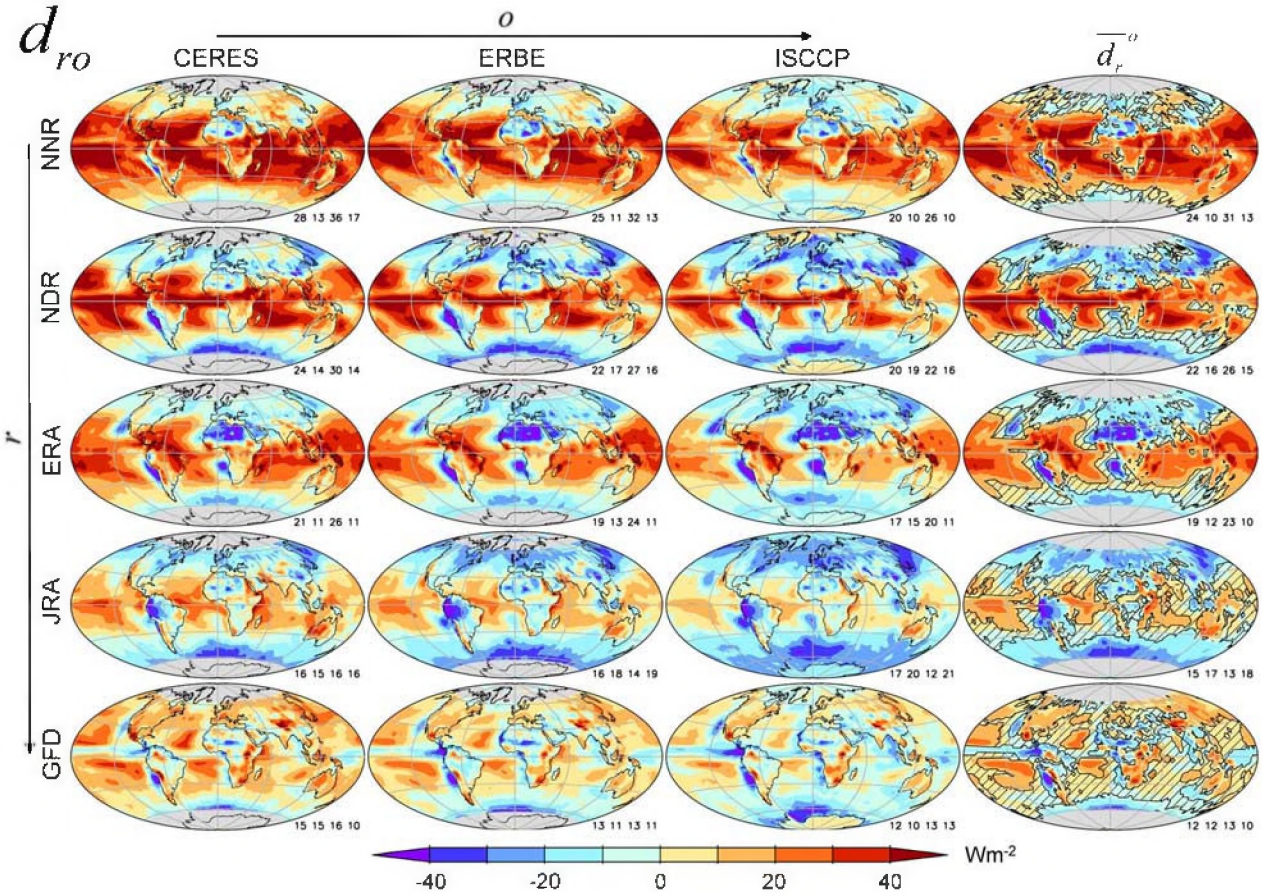

Figure 2. Differences (d_{ro}) in annual mean RSUT climatologies between the five products (NNR, NDR, ERA, JRA, and GFD) and three different observations (CERES, ERBE, ISCCP). Last column ($\overline{d_r}^o$) shows the mean difference over all three observations; hatching indicates insignificant (t test at 5% error level) differences. Numbers in lower right corners are RMS errors e (GL, NH, TR, and SH). Grey shading indicates insufficient data. Units are W m^{-2}.

observational uncertainties are relatively small and that the observations can be used to validate the reanalyses.

[21] Figure 2 (bottom row) shows the comparison of the model (GFD) against the observations. The error patterns of model and reanalyses share some common features, for example the negative biases over the upwelling regions and positive biases over the subtropical oceans. This leads us to speculate that perhaps similar deficiencies in the reanalysis assimilation models and the climate model are responsible for these errors. The magnitude of these errors, however, is generally smaller in the model ($\sim$14 W m^{-2} RMS error) than in the reanalyses. It is likely that the good performance of the model is due to tuning of the model's parameterization toward observed RSUT values or closely related fields.

[22] We also investigated how sensitive our results for RSUT were with respect to the particular years included in calculating the climatologies. In addition to 1979–1999, we took reanalysis climatologies over the ERBE period 1985–1989 (not shown). Keeping the observational base periods unchanged (Table 3), we found that the results were virtually unchanged using the new, much shorter time period used to form the reanalysis climatologies. The change in global RMS error amounted to less than $\pm$1 W m^{-2}. We therefore conclude that at least for this quantity, sampling uncertainty related to the specific choice of years is small compared to observational uncertainties and reanalysis errors.

5. Systematic Reanalysis Errors

[23] As noted before, the general RSUT error patterns exhibited by the reanalyses are largely independent from the choice of validating observations, indicating that these errors are real. Moreover, the patterns amongst the different reanalyses are also similar. This becomes more evident in the right column of Figure 2, which show for each product the average difference associated with the three observations. This is identical to taking the difference between one reanalysis and the mean over all observations (i.e., $\overline{d}^o = \overline{r - o}^o = r - \overline{o}^o$). In order to isolate robust from nonrobust features, we performed a t test at the 95% confidence level using the three difference maps as independent samples with two degrees of freedom. As indicated by the absence of hatching, most of the error patterns discussed before are statistically significant across all reanalyses. This suggests that reanalyses exhibit common systematic errors for the quantity RSUT.

[24] It is now of interest to find out whether the reanalyses show similar common biases in quantities other than RSUT. If such biases were found, this would create a situation similar to the common systematic bias problems known in the development of climate models, such as the Pacific cold-tongue [*Karnauskas et al.*, 2007], double intertropical convergence zone [*Mechoso et al.*, 1995], and cold biased stratosphere [*Cordero and Forster*, 2006] phenomena. Since reanalyses are routinely used for the validation of climate model output, we investigate the nature of possible systematic errors in the reanalyses.

[25] We define a systematic error in a particular quantity as the mean error over all reanalyses when validating against the mean of all available observations, i.e., $\overline{d}^o = \overline{r - \overline{o}^o}^r = \overline{r}^r - \overline{o}^o$ for $r = 1 \ldots 4$ (i.e., excluding GFD). Similarly as before, we tested the robustness of the results by performing a t test at the 95% level, now using the errors from the four different reanalyses as independent samples. This test strategy is sensitive to errors that occur in most reanalyses with respect to the mean observations. To the extent that the mean observations reflect the real state of the atmosphere, those differences can be classified as systematic reanalysis errors. We calculated the annual mean differences for all 18 "physics" quantities, the result of which is shown in Figure 3. In order to emphasize systematic errors, nonsignificant differences were removed and replaced by white shading.

[26] Figure 3 demonstrates the existence of systematic biases in almost all tested quantities. For example, over most of the extratropics the reanalyses have too little total cloudiness (CLT). Considering that clouds are simulated in the first generation products (NNR and NDR) by a diagnostic scheme and in the newer products (ERA and JRA) by a prognostic scheme, CLT biases that are common to all four products are somewhat surprising. As one would expect from the important control of clouds on radiation, these errors are to some extent accompanied by positive biases in shortwave cloud radiative forcing (CFST), outgoing longwave radiation at TOA (RLUT), and downwelling shortwave radiation at the surface (RSDS). (The spatial coherence amongst the different error patterns becomes more evident when errors that do not pass the t test at 95% are also considered.) On the other hand, there is some indication that the negative extratropical cloudiness biases are associated with same signed biases in precipitation (PR), outgoing shortwave radiation at TOA (RSUT), and downwelling longwave radiation at the surface (RLDS). Reanalyses also share a certain tendency of too much cloudiness in the tropics. The corresponding anomaly clusters are accompanied by physically consistent anomalies in PR and RSUT (both positive) and CFST and RSDS (both negative).

[27] We note that the existence of physically consistent biases between cloudiness and various radiative quantities is to some extent expected, since the majority of the validating observations for cloudiness and radiation were collected by the same observation systems (CERES and ISCCP). On the other hand, the good agreement between the patterns of cloudiness and precipitation biases, which were derived from independent observation systems, suggest that most of the above errors are real.

[28] Other prominent reanalyses biases are excessive upward longwave radiation (RLUS) over oceans, which is consistent with too warm surface air temperatures (TAS), and negative biases in sea level pressure (PSL) over many parts of the oceans. Reanalyses also seem to overestimate both latent (HFLS) and sensible (HFSS) heat fluxes over the Kuroshio current, the Gulf stream, and other parts of the oceans.

6. Normalized RMS Errors

[29] We now study the individual reanalysis products and their ability to match the observations. Because of space limitations, we limit this discussion to mean errors averaged over a few large regions (GL, NH, TR, SH). Our basic measure of error is the normalized RMS error E, introduced in section 3. This nondimensional error measure is useful

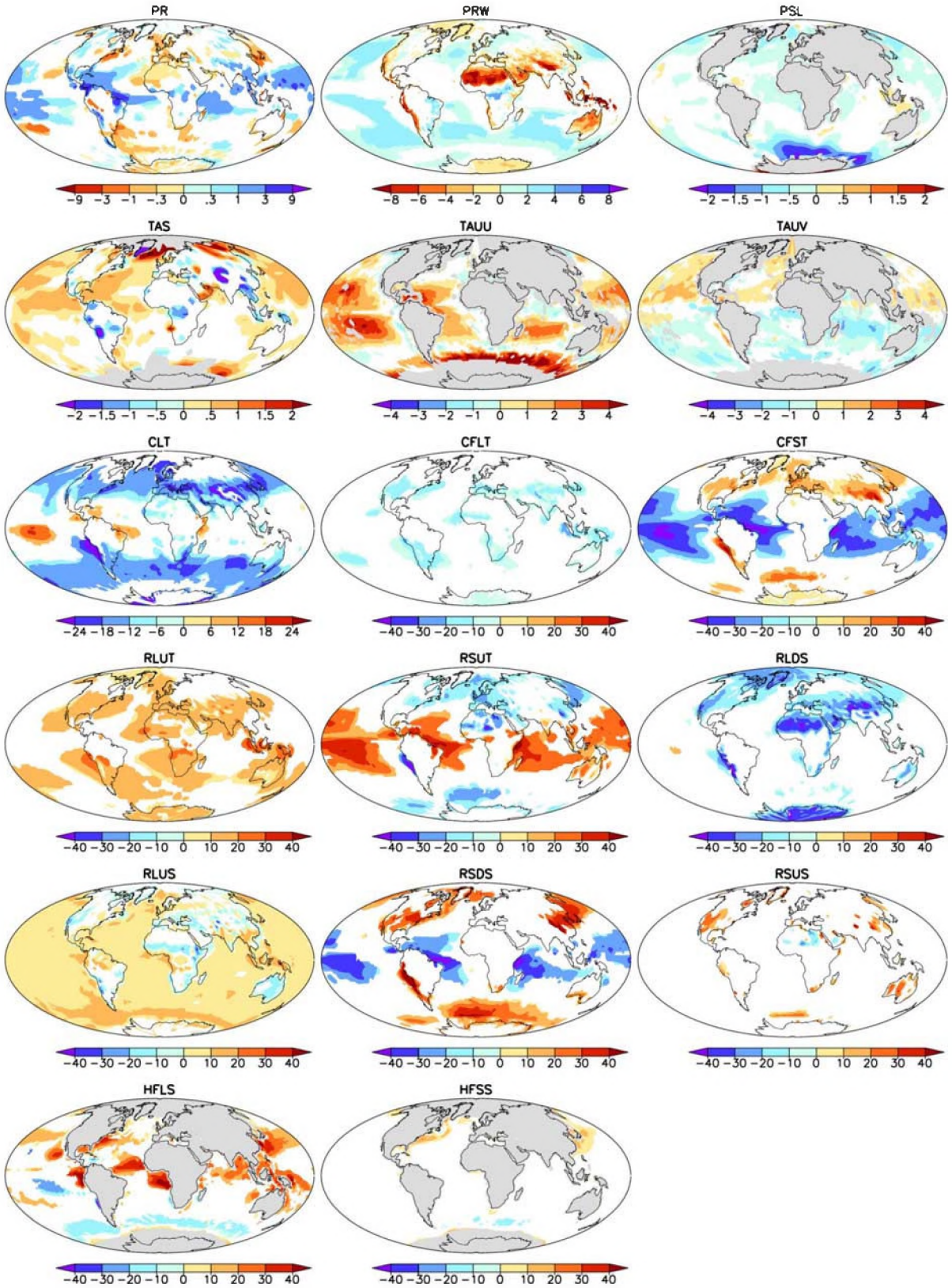

Figure 3. Annual mean common systematic biases $\overline{\overline{d}}^{IO}$ for the 18 "physics" quantities. White shading indicates nonsignificant (t test at 5% error level) errors, and grey shading indicates insufficient data. Because of poor spatial coverage, ICOADS data for TAS, HFLS, and HFSS were excluded from this analysis. See Table 2 for units.

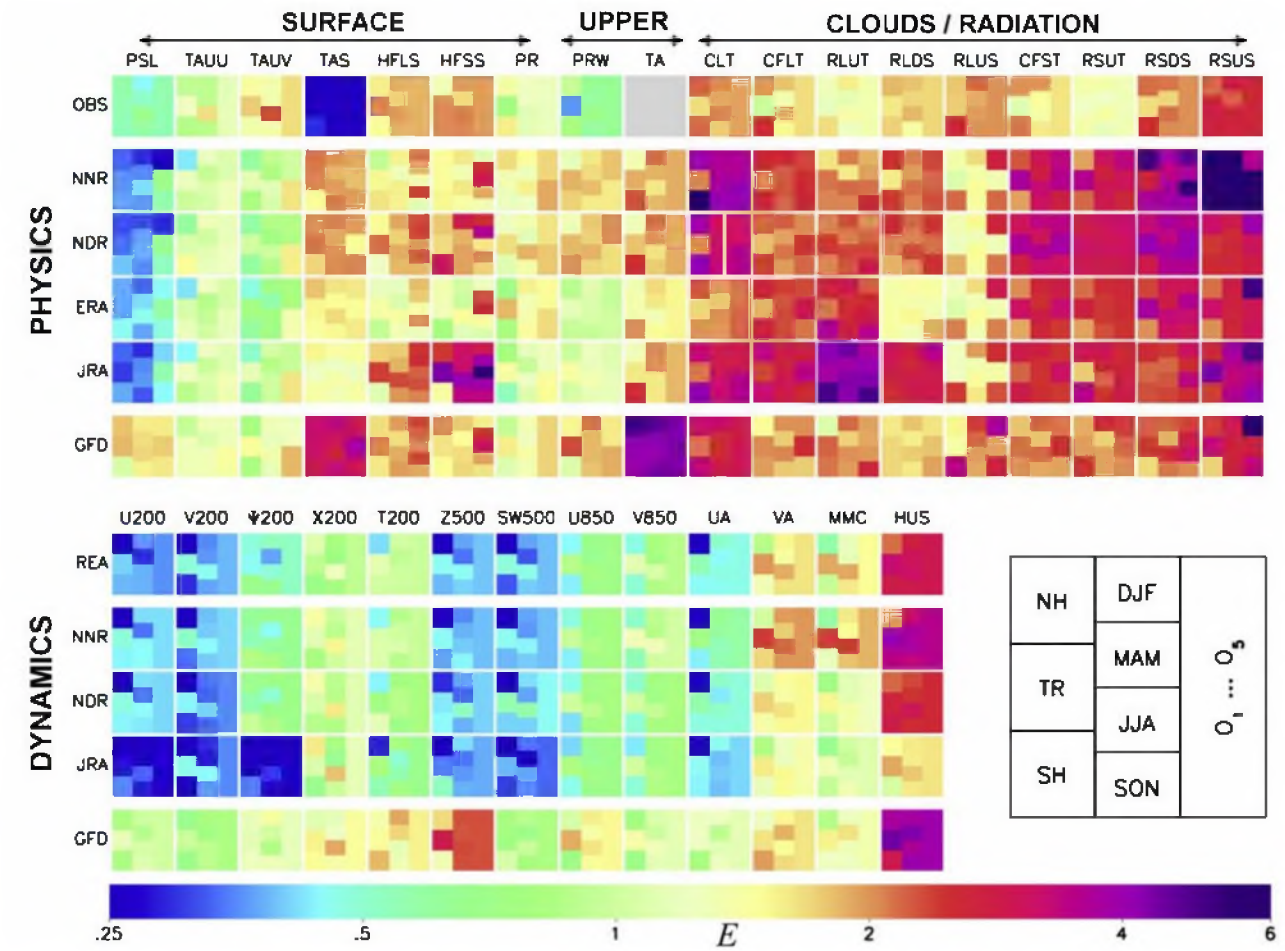

Figure 4. Normalized RMS errors E for (top) "physics" and (bottom) "dynamics" quantities. As indicated by legend, errors in each field are broken down by region n (averages over all seasons), season s (averages over all regions), and validating observations o (averages over all seasons and regions). Total number of validating observations varies by quantity (Table 2). Top rows in each group, entitled "OBS" and "REA," respectively, show E measured amongst multiple observations or reanalyses to provide a measure of observational uncertainty (not for TA since only single observation was available).

when comparing results from different quantities and regions since all errors are expressed in terms of their individual, local, interannual variabilities. Figure 4 illustrates the outcome of the intercomparison for the two groups of quantities: "physics" (Figure 4, top) and "dynamics" (Figure 4, bottom). Each field shows color coded values of E for a specific variable and product, broken down by region (annual and observational mean), season (global and observational mean), and validating observations (annual and global mean). The errors cover more than one order of magnitude (0.25 to 6), demonstrating that the ability of the reanalyses to reproduce the climatologies of the observations varies widely. In Table S4 of the auxiliary material[1], the interested reader can find the association between the fields O1–O5 and the validating observations.

6.1. Physics

[30] The results for the "physics" quantities are further grouped into variables belonging to the "surface," to the "free atmosphere," and to "clouds and radiation." The top row, labeled "OBS," shows E when validating multiple observations for the same quantity against each other. (This quantity could only be calculated at grid cells with multiple observations. Because of poor spatial coverage, we therefore excluded data from BSRN and GEBA from this type of calculation. For TA (zonal mean temperature), only one validating observational data set was available and thus it was not possible to derive an uncertainty measure.) It thus represents a normalized measure of observational uncertainty that can be directly compared with the errors for the reanalyses shown in the columns below. The general rule in interpreting these results is that the observational uncertainties (OBS) are tolerable as long as they are smaller than the E values of the reanalyses. For example, the observational uncertainties relative to the reanalysis error values are very small for TAS, and large for TAUx, HFxS, PR, and RLDS (see Table 2). In addition to E, we document in Table 4 the values of observational uncertainty in terms of the perhaps more common RMS error e.

[31] The "clouds and radiation" group of variables exhibits the largest errors, as shown by the predominance

[1]Auxiliary materials are available in the HTML. doi:10.1029/2007JD009278.

Table 4. Regional Breakdown of RMS Errors e Amongst Different Observations[a]

Variable	GL	NH	TR	SH	Units
PR	0.82	0.55	0.99	0.65	mm/d
PRW	1.0	1.5	0.8	1.1	mm
CLT	10	11	10	9.0	%
CFLT	5.7	6.4	5.2	6.4	W m^{-2}
CFST	10	11	8.8	11	W m^{-2}
RLUT	6.0	4.9	6.8	5.4	W m^{-2}
RSUT	9.7	9.2	11	7.8	W m^{-2}
RLDS	13	15	12	12	W m^{-2}
RLUS	14	14	15	14	W m^{-2}
RSDS	17	16	14	21	W m^{-2}
RSUS	14	13	7.4	21	W m^{-2}
PSL	0.76	0.94	0.40	0.14	hPa
TAS	0.12	0.17	0.10	0.13	K
TAUU	1.6	1.8	1.6	1.7	10^{-2} N m^{-2}
TAUV	1.7	1.7	1.7	1.8	10^{-2} N m^{-2}
HFLS	20	16	24	13	W m^{-2}
HFSS	6.8	7.2	4.7	9.6	W m^{-2}

[a]Shown are averages over the four seasons. Because of poor spatial coverage, ICOADS (only for TAS, HFLS, and HFSS), GEBA, and BSRN data were excluded from this analysis.

of warm, reddish colors. Most of the "clouds and radiation" quantities are rather well observable from space or ground, leading to relatively small observational uncertainties relative to the reanalyses. Exceptions are the surface longwave fluxes (RLxS), which for some of the observations (ISCCP, CERES) are highly derived from radiative transfer codes. Some of the largest errors of the reanalyses are associated with total cloudiness (CLT). Owing to the importance of clouds for the optical properties of the atmosphere, the CLT problem can explain much of the errors seen in the remaining radiative quantities, in particular those related to shortwave radiation at the top of the atmosphere (RSUT, CFST) and at the surface (RSxS). The errors seen in CLT may to some extent also be related to the lack of a commonly accepted definition of the critical optical thickness for a grid point to become a cloud. While the definition is reasonably close between the two validating data sets (CERES and ISCCP), it is unclear how it varies amongst the different reanalyses and the model, and to what extent this affects the outcome of our comparison. Note that the large error seen in NNR for shortwave radiation reflected at the surface (RSUS) is a well-documented problem [*Kanamitsu et al.*, 2002; *Weare*, 1997].

[32] In contrast to "clouds and radiation," "surface" variables exhibit much smaller errors. For example, sea level pressure (PSL) is very well reproduced by the reanalyses. This, of course, is expected, since most available oceanic PSL observations are assimilated in more or less the same way into all the reanalyses. This and the smoothly varying character of this field lead to an excellent representation in the reanalyses.

[33] The situation is quite different for surface air temperature (TAS), which is also a well observed quantity with small observational uncertainties. Compared to this, the reanalyses errors associated with TAS are large. This is somewhat surprising, since for over 70% of the globe, surface air temperatures are tightly coupled to sea surface temperatures (SSTs), and since SSTs are prescribed from observations to the reanalyses. More detailed analysis (not shown) indicates that the TAS errors occur mostly over land, where there is no beneficial effect of prescribing perfect SSTs. Over land, one also has to keep in mind that TAS depends on the height of the observations. Depending on how realistic topography is represented in the different models, certain discrepancies with respect to TAS observations are to be expected.

[34] Overall, there exist important differences in how faithful the different reanalyses reproduce observed climate for the "physics" quantities. Most notably, the second generation reanalysis ERA performs consistently well, and in most quantities better than any other product. The only major exception from this rule is tropical precipitation (PR), where ERA performs worse than any other product. Even the model (GFD) produced precipitation is more close to the observations than ERA. Further inspection (see also Tables S5–S8) reveals that ERA consistently produces excessive precipitation over the tropics, a major deficiency which has already been documented before [*Troccoli and Kallberg*, 2004].

[35] For JRA, the most recent reanalysis, the situation is inconclusive. Comparing JRA to the other products, the errors in TAS and PR are small, but considerable biases exist for some other quantities, such as HFSS, RLUT, and RLDS.

[36] Figure 4 also illustrates the performance of the model (GFD). For some quantities, such as surface temperature (TAS), air temperature (TA), and sea level pressure (PSL), the model exhibits much larger biases than the reanalyses. For most other quantities, however, the model is as good as or even better than the reanalyses. This is unexpected, considering that the reanalyses are strongly constrained by observations, whereas the (coupled) model is not. The good performance of the model compared to the reanalyses is most evident in radiative quantities. Here, reanalyses are probably least constrained by observations. On the other hand, tuning toward observed climatologies, as well as modern parameterization packages are the likely reasons why the model achieves such realistic climatologies.

6.2. Dynamics

[37] Figure 4 (bottom) shows the errors in reproducing the "dynamics" group of quantities. This group was validated against ERA since observations were unavailable. Similarly to before, the top row, now labeled "REA," represent a measure of uncertainty. It shows E when comparing the four reanalyses against each other.

[38] The errors and uncertainties in the "dynamics" quantities are much smaller than those in the "physics" quantities seen before. At least three reasons may help to explain this reduction. First, reanalyses rely on similar upper air observations. Although the details in how this information is assimilated into the reanalyses differ from product to product, this should help to make the "dynamics" of the reanalyses similar to each other. Second, dynamical quantities of the free atmosphere exhibit smaller spatial and temporal gradients than the "physics" quantities, making it is easier to reproduce these fields. Third, ERA data are used for verification, with the consequence that systematic errors in the "dynamics" quantities remain undetected. As expected, the model generally does not perform as well as the reanalyses, except for MMC and VA.

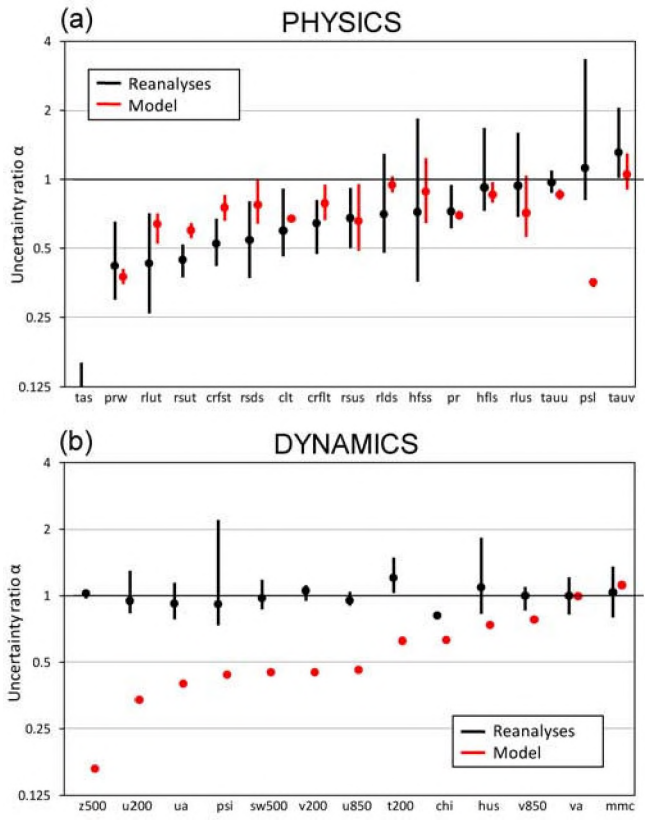

Figure 5. Uncertainty ratio α for (a) "physics" and (b) "dynamics" quantities derived from (black) reanalyses and (red) the model. Circles denote average α, and vertical line shows the range of outcomes using individual reanalysis, model, and reference data. All quantities are for the global domain and averaged over the four seasons.

[39] Specific humidity (HUS) is the only quantity which exhibits considerable errors ($E \sim 4$) within this group of quantities. We originally planned to validate HUS against data from the AIRS instrument [*Chahine et al.*, 2006], but later we decided against it because of large and inconsistent discrepancies between data from AIRS and the reanalyses, casting doubts on the validity of AIRS data as reference for HUS. We note as possible explanation the relatively short time period of AIRS (2002–2006), making the calculation of dependable interannual variability difficult. Moreover, the AIRS instrument is unable to measure through clouds, introducing additional biases into the HUS estimates. Despite those difficulties, it is probably useful to mention that GFD and ERA exhibit smallest and NNR and NDR largest HUS errors ($E \sim 6$) when validating against AIRS. Spatially, all five products seem to reproduce the AIRS observations better in the Tropics than in the extratropics.

[40] Tables S1–S3 of the auxiliary material provide for all quantities tabulated values of the RMS errors e and the observed interannual variabilities. In addition, Tables S5–S8 of the auxiliary material give numerical values of the mean biases b, stratified by region, validating observation, and season. Consistency or inconsistency of the sign and magnitude of b across the different observations (for the same product) and across different products (for the same observation) gives additional clues of how realistic the diagnosed uncertainties and errors are.

7. Uncertainty Ratios

[41] We now investigate quantitatively the magnitude of the observational uncertainties in comparison to typical reanalyses and model errors E. As mentioned before, the observational uncertainties are tolerable as long as they are smaller than the errors of the data to be verified, or in other words, as long as the ratio between observational uncertainty and reanalysis error is less than unity. We call this number the uncertainty ratio, and denote it α.

[42] Figure 5a shows the average α and its range of values for the "physics" quantities. Here, α is defined as the ratio between the average E amongst different observations and the average E from the reanalyses/model validated against the observations. The range, which is the result from pairing different reanalyses or the model with different validating observations, is an indicator of how similar the uncertainties between the various products are. In more mathematical terms, the black signatures represent the mean and the range of $E_{OBS}/E_{\vec{REA}\,vs.\,\vec{OBS}}$, where the three dots indicate which quantity was varied to calculate the range of outcomes. Similarly, the red signatures represent $E_{OBS}/E_{GFD\,vs.\,\vec{OBS}}$.

[43] For the reanalyses (black signatures), the uncertainty ratio is in most cases smaller than one. Notable exceptions are quantities located to the very right of the graph, such as sea level pressure (PSL), as well as radiative (RLUS) and nonradiative (HFxS, TAUx) surface flux quantities. The rather high uncertainty ratio for sea level pressure is expected, since it is a well observed quantity that is assimilated in similar ways into all the reanalyses. The situation for the surface fluxes is just the opposite: these quantities are usually not directly observable, leading to high uncertainties in their observational estimates.

[44] For the model (red signatures), the ratios tend to be somewhat larger (except PSL), which is just another sign that the model is performing often as well as or even better than some of the reanalyses.

[45] Figure 5b shows the uncertainty ratios for the "dynamics" quantities. Since in this case no observations were available, we calculated α from "reanalysis uncertainties" (from comparing different reanalyses against each other) and from "reanalysis errors" (when validating against ERA) (both shown in Figure 4), i.e., $\alpha = E_{REA}/E_{\vec{REA}\,vs.\,ERA}$. Now, α for the reanalyses (black signatures) is close to one, simply indicating that the reanalyses are similar to each other. The ratios for the model (red circles), which were calculated according to $\alpha = E_{REA}/E_{GFD\,vs.\,ERA}$, are much smaller than one, except for the meridional winds (VA, V850) and the closely related meridional mass stream function (MMC). Note that in this case no error bars are displayed since only one model realization was available. We conclude that for most "dynamics" quantities the reanalyses agree more with each other than with the model. Since the "dynamics" of the reanalyses are strongly constrained by upper air observations, it is reasonable to assume that reanalyses are more reliable than model data. Therefore, it is

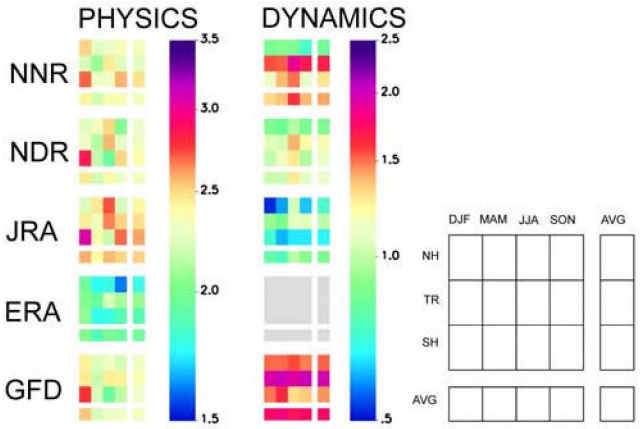

Figure 6. Normalized RMS errors E when validating against the mean of all available observation, averaged over all (left) "physics" and (right) "dynamics" quantities. As indicated by legend, errors are shown by individual region n, season s, along with their respective averages. Fields in lower right corners are global and annual averages.

in most cases safe to validate climate model output against reanalyses.

8. Combined Errors

[46] We finish this paper by providing a comparison of the combined errors seen in the climate mean states of the different reanalyses and the model. To this end, we average the normalized RMS errors from each product across the different climate quantities. Arguably, there is no clear physical meaning associated with the average error across different climate quantities, but this procedure conveniently provides a first-order estimate of overall error. We justify our approach by utilizing normalized error quantities when averaging, which ensures that each quantity is weighted roughly equally in the final result. We also note that a very similar approach has been used by us [*Reichler and Kim*, 2008] and others [*Murphy et al.*, 2004] in the context of climate model validation.

[47] Figure 6 shows the combined reanalysis and model errors, which are again separated into "physics" and "dynamics" quantities. In order to be better able to emphasize differences, the color schemes were modified with respect to Figure 4. Where applicable, the mean over all available observations was used for the validation. As indicated by the legend, the fields for each product are subdivided into errors for the four seasons, the three regions, and averages thereof. This shows how the combined errors vary in space and time of the year.

[48] With respect to the "physics" quantities, NNR, NDR, and JRA share some important features. For example, the combined errors in all three reanalyses are largest over the SH, in particular during spring (SON) and summer (DJF). We assume that despite the assimilation of spaceborne observations into the reanalyses, the scarcity of ground-based observations is responsible for the relatively large errors over this part of the world. The model (GFD) also exhibits largest combined errors over the SH during summer. Over the NH, NDR and JRA also show a tendency of largest errors during spring (MAM) and summer (JJA). It is not exactly clear why the errors during the "quiet" seasons tend to be largest, but we speculate that during this time of the year the atmosphere is less governed by simple geostrophic balance. Rather, other more complicated effects like convective activity become more important in determining the state of the atmosphere. Interestingly, ERA performs consistently well during all seasons and over all regions.

[49] For the "dynamics" quantities, JRA is most similar to ERA, whereas the model shows largest discrepancies with respect to ERA. In the tropics, NNR exhibits rather larger differences with respect to ERA. NDR and JRA do not have such large biases in the tropics, suggesting that the differences of NNR in the tropics are real errors.

9. Summary and Conclusions

[50] The present study provided a systematic assessment of uncertainties contained in the climate mean state of global observations, four reanalysis products, and one state-of-the-art climate model. The study was guided by the need to better understand the reliability of global climate data sets in the context of climate model validation. The investigation was based on 18 "physics" quantities, for which in most cases, multiple observations were available. Additionally, we included 13 "dynamics" quantities, using ERA reanalysis as a reference.

[51] In a manner similar to climate models, reanalyses exhibit significant systematic errors in almost all climate quantities investigated. We speculate that similar uncertainties in the formulation of the different assimilation models are the underlying cause for such systematic errors. By breaking normalized root mean square errors down by season, region, and validating observations, we demonstrated that the ability of reanalysis to reproduce the observed climate mean state is a strong function of the individual product and the quantity under consideration. The largest errors were found in radiative quantities, and often the coupled model output was closer to the observations than some of the reanalyses.

[52] Usually, the observational uncertainties were smaller than the reanalyses or model errors. However, for some surface flux quantities the uncertainties were as large as the errors themselves. In this case, the results of the comparison are ambiguous: Either the reanalyses perform particularly well or (some of) the observations are poor. Overall, the climate mean state produced by the ERA reanalysis matched best the available observations. For dynamical quantities of the free atmosphere, the discrepancies amongst the reanalyses were usually much smaller than the differences between model and reanalyses. Notable exceptions were meridional circulation quantities and specific humidity.

[53] One of the main challenges of this study was the lack of accurate validating observations. We tried to ameliorate this problem by using multiple observations and by quantifying observational uncertainty. However, there were additional problems, like the often inadequate global and/or temporal coverage of the observations, and the uncertainty to what extent the observations have already been assimi-

lated into the reanalyses. Another caveat of this study was the somewhat arbitrary choice of climate quantities, which were not selected by physical reasoning, but mandated by the available observations. One of the consequences was that the error information contained in the different climate quantities was to some extent redundant. An example of this can be seen from the similarity between the results for zonal mean meridional wind and the meridional mass stream function. We last note that this study was solely focused on the climate mean state. Mean climate, however, is just one particular aspect of climate, and variability and long-term trends are at least as important in the context of climate research.

[54] We were somewhat surprised by the good performance of the climate model (GFD) in comparison to the reanalyses. We found in a different study [*Reichler and Kim*, 2008] that this particular climate model simulates a very realistic climate in comparison to other models, but the present study showed that this model sometimes even surpasses the quality of reanalyses. Since further model improvements are expected in the future, this seems to suggest that it will soon be difficult to find appropriate validation data for models. As we have seen, this problem is already real with respect to surface fluxes and some other, dynamical quantities. In this study, we tried to deal with this problem by averaging across multiple data sets for the same quantity. The rationale behind this is that the averaging procedure filters the errors to the extent that they are randomly distributed about the truth.

[55] In summary, the uncertainties of reanalyses and observations are sometimes large with respect to modern climate models, leading in extreme cases to situations where a meaningful validation of model output cannot be accomplished. This underlines the need for better climate observations, and it raises the expectation for a future reanalysis project with improved performance.

[56] **Acknowledgments.** We thank Paul Staten for providing useful comments on an earlier version of this paper. We also thank the three anonymous reviewers for their constructive comments. We acknowledge GFDL for providing the model output and the many other groups and centers for providing observation-based data. This work was supported by grants from NSF (ATM0532280) and NOAA (NA06OAR4310148).

References

Adler, R. F., et al. (2003), The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), *J. Hydrometeorol.*, *4*(6), 1147–1167.

Allan, R., and T. Ansell (2006), A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004, *J. Clim.*, *19*(22), 5816–5842.

Barkstrom, B. R. (1984), The Earth Radiation Budget Experiment (ERBE), *Bull. Am. Meteorol. Soc.*, *65*(11), 1170–1185.

Bengtsson, L., S. Hagemann, and K. I. Hodges (2004), Can climate trends be calculated from reanalysis data?, *J. Geophys. Res.*, *109*, D11111, doi:10.1029/2004JD004536.

Betts, A. K., M. Zhao, P. A. Dirmeyer, and A. C. M. Beljaars (2006), Comparison of ERA40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets, *J. Geophys. Res.*, *111*, D22S04, doi:10.1029/2006JD007174.

Bromwich, D. H., and R. L. Fogt (2004), Strong trends in the skill of the ERA-40 and NCEP-NCAR reanalyses in the high and midlatitudes of the Southern Hemisphere, 1958–2001, *J. Clim.*, *17*(23), 4603–4619.

Chahine, M. T., et al. (2006), AIRS: Improving weather forecasting and providing new data on greenhouse gases, *Bull. Am. Meteorol. Soc.*, *87*(7), 911–926.

Chou, S.-H., E. Nelkin, J. Ardizzone, R. M. Atlas, and C.-L. Shie (2003), Surface turbulent heat and momentum fluxes over global oceans based on the goddard satellite retrievals, version 2 (GSSTF2), *J. Clim.*, *16*(20), 3256–3273.

Cordero, E., and P. M. D. F. Forster (2006), Stratospheric variability and trends in models used for the IPCC AR4, *Atmos. Chem. and Phys.*, *6*, 5369–5380.

Delworth, T. L., et al. (2006), GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics, *J. Clim.*, *19*(5), 643–674.

Gibson, J. K., P. Kallberg, S. Uppala, A. Nomura, A. Hernandez, and E. Serrano (1997), ERA description, *ECMWF ERA-15 Proj. Rep. Ser.*, *1*, Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K. (Available at http://www.ecmwf.int/publications)

Gilgen, H., and A. Ohmura (1999), The global energy balance archive, *Bull. Am. Meteorol. Soc.*, *80*(5), 831–850.

Grassl, H., V. Jost, R. Kumar, J. Schulz, P. Bauer, and P. Schluessel (2000), The Hamburg Ocean-Atmosphere Parameters and Fluxes From Satellite Data (HOAPS): A climatological atlas of Satellite-derived air-sea-interaction parameters over the oceans, Max Planck Inst. for Meteorol., Hamburg, Germany.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor (1999), Surface air temperature and its changes over the past 150 years, *Rev. Geophys.*, *37*, 173–199.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*(3), 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter (2002), NCEP-DOE AMIP-II reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, *83*(11), 1631–1643.

Karnauskas, K. R., R. Murtugudde, and A. J. Busalacchi (2007), The effect of Galapagos Islands on the equatorial Pacific cold tongue, *J. Phys. Oceanogr.*, *37*, 1266–1281.

Kubota, M., N. Iwasaka, S. Kizu, M. Konda, and K. Kutsuwada (2002), Japanese Ocean Flux Data Sets with Use of Remote Sensing Observations (J-OFURO), *J. Phys. Oceanogr.*, *58*(1), 213–225.

Liebmann, B., and C. A. Smith (1996), Description of a complete (interpolated) outgoing longwave radiation dataset, *Bull. Am. Meteorol. Soc.*, *77*, 1275–1277.

Lu, J., G. A. Vecchi, and T. Reichler (2007), Expansion of the Hadley cell under global warming, *Geophys. Res. Lett.*, *34*, L06805, doi:10.1029/2006GL028443.

Marshall, G. J. (2002), Trends in Antarctic geopotential height and temperature: A comparison between radiosonde and NCEP-NCAR reanalysis data, *J. Clim.*, *15*(6), 659–674.

Marshall, G. J. (2003), Trends in the Southern Annular Mode from observations and reanalyses, *J. Clim.*, *16*(24), 4134–4143.

Mechoso, C. R., et al. (1995), The seasonal cycle over the tropical Pacific in coupled ocean–atmosphere general circulation models, *Mon. Weather Rev.*, *123*(9), 2825–2838.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth (2004), Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, *430*, 768–772.

Newman, M., P. D. Sardeshmukh, and J. W. Bergman (2000), An assessment of the NCEP, NASA, and ECMWF reanalyses over the tropical West Pacific warm pool, *Bull. Am. Meteorol. Soc.*, *81*(1), 41–48.

Ohmura, A., et al. (1998), Baseline Surface Radiation Network (BSRN/WRMC), a new precision radiometry for climate research, *Bull. Am. Meteorol. Soc.*, *79*, 2115–2136.

Onogi, K., et al. (2007), The JRA-25 reanalysis, *J. Meteorol. Soc. Jpn.*, *85*(3), 369–432.

Randel, D. L., T. J. Greenwald, T. H. Vonder Haar, G. L. Stephens, M. A. Ringerud, and C. L. Combs (1996), A new global water vapor dataset, *Bull. Am. Meteorol. Soc.*, *77*(6), 1233–1246.

Reichler, T., and J. Kim (2008), How well do coupled models simulate today's climate?, *Bull. Am. Meteorol. Soc.*, in press.

Renwick, J. A. (2004), Trends in the Southern Hemisphere polar vortex in NCEP and ECMWF reanalyses, *Geophys. Res. Lett.*, *31*, L07209, doi:10.1029/2003GL019302.

Rossow, W. B., A. W. Walker, D. E. Beuschel, and M. D. Roiter (1996), International Satellite Cloud Climatology Project (ISCCP), Documentation of new cloud datasets, *WMO/TD-737*, 115 pp., World Meteorol. Organ., Geneva, Switzerland.

Smith, T. M., and R. W. Reynolds (2004), Reconstruction of monthly mean oceanic sea level pressure based on COADS and station data (1854–1997), *J. Atmos. Oceanic Technol.*, *21*, 1272–1282.

Smith, T. M., and R. W. Reynolds (2005), A global merged land-air-sea surface temperature reconstruction based on historical observations (1880–1997), *J. Clim.*, *18*(12), 2021–2036.

Sterl, A. (2004), On the (in)homogeneity of reanalysis products, *J. Clim.*, *17*(19), 3866–3873.

Trenberth, K. E., D. P. Stepaniak, J. W. Hurrell, and M. Fiorino (2001), Quality of reanalyses in the tropics, *J. Clim.*, *14*(7), 1499–1510.

Troccoli, A. and P. Kallberg (2004), Precipitation correction in the ERA-40 reanalysis, *ERA-40 Proj. Rep. Ser.*, *13*, Eur. Cent. for Med.-Range Weather Forecasts, Reading, U. K.

Uppala, S. M., et al. (2005), The ERA-40 reanalysis, *Q. J. R. Meteorol. Soc.*, *131*(612), 2961–3012.

Watterson, I. G. (1996), Non-dimensional measure of climate model performance, *Int. J. Climatol.*, *16*, 379–391.

Watterson, I. G., and M. R. Dix (2005), Effective sensitivity and heat capacity in the response of climate models to greenhouse gas and aerosol forcings, *Q. J. R. Meteorol. Soc.*, *131*, 259–279.

Watterson, I. G., M. R. Dix, and R. A. Colman (1999), A comparison of present and doubled CO2 climates and feedbacks simulated by three GCMs, *J. Geophys. Res.*, *104*, 1943–1956.

Weare, B. C. (1997), Comparison of NCEP-NCAR cloud radiative forcing reanalyses with observations, *J. Clim.*, *10*, 2200–2209.

Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee III, G. L. Smith, and J. E. Cooper (1996), Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System experiment, *Bull. Am. Meteorol. Soc.*, *77*(5), 853–868.

Woodruff, S. D., R. J. Slutz, R. L. Jenne, and P. M. Steurer (1987), A comprehensive ocean-atmosphere data set, *Bull. Am. Meteorol. Soc.*, *68*(10), 1239–1250.

Xie, P. P., and P. A. Arkin (1996), Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, *J. Clim.*, *9*(4), 840–858.

Yu, L., and R. A. Weller (2007), Objectively Analyzed air-sea heat fluxes for the global ice-free oceans (1981–2005), *Bull. Am. Meteorol. Soc.*, *88*(4), 527–539.

Zhang, Y., W. B. Rossow, A. A. Lacis, V. Oinas, and M. I. Mishchenko (2004), Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data, *J. Geophys. Res.*, *109*, D19105, doi:10.1029/2003JD004457.

J. Kim and T. Reichler, Department of Meteorology, University of Utah, 135 S 1460 E, Room 819 (WBB), Salt Lake City, UT 84112-0110, USA. (thomas.reichler@utah.edu)